DOCUMENT RESUME

ED 218 320 TM 820 360

AUTHOR	Cabello, Beverly
TITLE	Potential Sources of Bias in Dual Language Achievement Tests.
INSTITUTION	California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
REPORT NO	CSE-R-178
PUB DATE	81
NOTE	28p.
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Achievement Tests; Bilingual Education; *Bilingualism; Culture Fair Tests; Elementary Education; English (Second Language); *Evaluation Needs; Spanish Speaking; Standardized Tests; *Test Bias; *Translation
IDENTIFIERS	*Comprehensive Tests of Basic Skills

ABSTRACT

Potential sources of bias in dual language achievement tests were identified and examined. Content, concepts, and vocabulary presented in monolingual English curricula may differ drastically from those presented in bilingual curricula. The Spanish or English versions of the Comprehensive Tests of Basic Skills (CTBS) were administered to 1259 students in 81 second, third, fifth and sixth grade classrooms in 5 California school districts. Statistical and content analysis provided evidence of three possible sources of bias: problems inherent in translation, the match between the test and instructional material, and intervening cultural values. Dual language tests also raise the question of whether, and how, standard achievement tests can match the wide variety of curricula used by American schools. Further empirical studies are necessary to isolate the cause of bias for individual test items, whether in the CTBS or any other test. Studies providing more information concerning biases stemming from cultural interference and curricular match are needed in order to develop a methodology beyond the statistical approaches now available. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED218320

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

POTENTIAL SOURCES OF BIAS IN DUAL LANGUAGE ACHIEVEMENT TESTS

Beverly Cabello

SCOPE OF INTEREST NOTICE

The ERIC Facility has assigned this document for processing to: TM FL

In our judgement this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.

CSE Report No. 178
1981

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

G. Gray

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 820 360

CENTER FOR THE STUDY OF EVALUATION
Graduate School of Education
University of California, Los Angeles

The project presented or reported herein was performed pursuant to a grant from the National Institute of Education, Department of Education. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

INTRODUCTION

In response to the instructional and evaluation needs of bilingual programs, many dual language tests and basal readers have been developed by both commercial publishers and public school educators. Most of these materials are designed to serve the Spanish-speaking sector of the student population because they are the largest language minority population receiving bilingual education (Comptroller General's Report, 1978). However, creating a Spanish language test which is equally comprehensible, useful, and fair to Mexicans, Puerto Ricans, and other Hispanic students of varying educational and social backgrounds, is as difficult as creating a test in English to serve American, British, Australian, and other English speaking students around the world equally well. A multitude of factors intervene, such as regional differences in vocabulary and social class differences in language and academic skills. However, the problem of creating a test to be equally fair to all Hispanic students is exacerbated by the fact that it will probably be a translation of a test originally developed for English speakers. Thus, in addition to the intervening factors mentioned above, the technical problems inherent to the art of translation must also be considered.

Translating a test written in "standard-middle-class-English," one which reflects many American values and behaviors, into another language which may not share these values or behaviors is a troublesome proposition for test developer and test consumer alike. There are three possible avenues to resolve this problem:

(1) the developer could attempt to translate from English to the target language in such a way as to create a "culture-free" test;

(2) the developer and/or consumer might identify the culture-laden items and attempt take them into account when interpreting the results of the translated test;

(3) a separate test in the second language could be developed so as to reflect the culture and curricula of the target population. None of these choices is an easy one. As mentioned above, most dual language tests are translations or adaptations of material which was originally written in English. Many are developed under the assumption that dual language materials are equivalent in content, difficulty, reliability, validity, and other features (cf CTBS manual). Often, consumers do not or cannot examine the translation for its accuracy, appropriateness for the target population, or technical qualities (BETA Needs Assessment, 1978). Some may think that the non-English version is written in a "standard" language and is therefore appropriate for use with any population speaking that language. Hence, materials which were developed for one population (e.g., Puerto Rican) are often used with another (e.g., Mexican) without going through an appropriate process of validation with the second group.

Psychometric properties do not carry over to the translated test. One basic difference between languages that disrupts "carry over" is variation in the frequency of word use and word difficulty. Words which may be commonplace and "easy" in one language (such as saddle, spaceship, or chocolate chip cookie) are not equally so in another language (Rodrigues, 1956). The use of dual language materials also implies that both the English and the limited-English population receive similar curricula, and therefore their performance on these tests can be compared. However, the content, concepts, and vocabulary presented in monolingual English curricula may differ drastically from those presented in bilingual curricula. In an effort to examine these assumptions about dual language Spanish/English achievement tests, the Center for the Study of Evaluation conducted studies in test bias. The purpose of this paper is to describe some of the sources of bias which were identified in ASCE study and to offer examples of them.

METHODS

Spanish and English versions of the California Tests of Basic Skills (CTBS) and its Spanish adaptation were selected for examination because CTBS is a widely used instrument. Also, at the time of instrument review and selection, it was the only commercially available standardized achieve-

TABLE 1

Item Numbers	Items which were easier in English				Items which were easier in Spanish			
	Vocabulary Level C	Vocabulary Level 2	Reading Comprehension Level C	Reading Comprehension Level 2	Vocabulary Level C	Vocabulary Level 2	Reading Comprehension Level C	Reading Comprehension Level 2
1		*	*	*				
3				*				
4			*					
5			*					
6		*	*		*			
7	*			*				
8		*						
9		*						
10								
11		*						
12	*	*						
13		*	*					
14	*		*					
15	*	*		*				
16	*							
17				*				
18				*				
19		*						
20	*	*						
21				*				
22				*				
23	*	*						
24				*				
25		*		*				
29					*			
30						*		
32					*			
34					*			
35		*						
36		*						
38				*				
39				*				
40						*		
41				*				

ment test in the two languages.

The CTBS and its Spanish counterpart were also selected because of the assumptions and procedures under which the two versions were developed. The intent of the publisher was to provide two tests which were as similar as possible in their rationale, and in their process/content classification scheme. The objectives measured by both language versions are:

1. the ability to recognize or recall information;
2. the ability to translate or convert concepts from one kind of language (verbal or symbolic) into another;
3. the ability to comprehend concepts and their interrelationships;
4. the ability to apply techniques, including performing operations;
5. the ability to extend interpretation beyond stated information (CTBS, 1974/78).

Both language versions of the test, at each test level, are the same in length, timing, and administration procedures. The tests were subjected to editorial procedures designed to reduce bias, and separate ethnic group pilot studies, as well as reviews for bias, were conducted. Although the translation was as close as possible, efforts to keep the test content and its measurement features intact sometimes prohibited such precision and adaptations were necessary.

Five school districts in California participated in the CSE study, and involved 1259 students in 81 classrooms (2nd and 3rd; and 5th and 6th grades). Each school had at least two classrooms at each grade level. Students had to be either in a Title VII bilingual program or in a monolingual English program. Pull out, ESL, and departmentalized programs were excluded. Student language proficiency was determined by the school districts and they elected to give students either the Spanish or English version of the CTBS based on teachers' judgement.

The CTBS English level C, designed for students in grades 1.6 to 2.9, and its Spanish counterpart, designed for grade 2, was administered to the second and third graders as part of regular school district test procedures. The fifth and sixth graders were tested under the same conditions. They received the CBTS English level 2, designed for grades 4.5 to 6.9, and the Spanish translation, designed for students in the fifth and sixth grades.

Second and third graders use selected for study because most bilingual programs in the United States consist of K-3 programs (Comptroller General Report - 1978). Furthermore, kindergarten and first grade students do not usually receive group achievement tests. In addition, the question of their primary oral language development would have confounded the issue of bias (Saville - Troike, 1975). In addition, second and third grade students are generally tested in groups, they participate in achievement testing as part of the schooling process, and they have mastered most oral language skills in their primary language (Cummins, 1979).

Fifth and sixth grade students were selected to allow us to examine whether there were any developmental differences affecting bias. Furthermore, it was assumed that fifth and sixth graders had more mastery of reading comprehension skills in their primary language than the younger children.

Analyses. The full study used statistical and content anlaysis to examine potential bias. Statistical analyses indicated that several items on these tests were biased (See Table 1).

The items identified as biased were scrutinized to locate potential sources of bias such as the quality of the translation, curricular relevance, and cultural interference.

The quality of translation. The translation was reviewed for its accuracy in terms of meaning, connotations, style, and degree of difficulty

of key vocabulary and passages of the biased items. Poor translation of any of these aspects can result in misleading or confusing language which can interfere with the student's ability to comprehend test items and answer questions about them. A translation or adaptation must reflect not only the meaning of the original item, but should also maintain the same intent, difficulty level, style, and tone; otherwise the item's basic construct may be changed.

Item examination revealed several "popular distractors;" that is, a form of bias which occurs when either the English- or the Spanish-speaking populations consistently performs much better or worse on an item than the other group, and selects a particular distraction over the correct response.

A single English word may have several alternative translations available in Spanish. In certain instances these alternatives reflect regional preferences and/or degrees of difficulty. Thus, a translation may appear to favor one Spanish-speaking population over another or may not retain a constant level of difficulty. In some cases, a seemingly correct translation may vary significantly in meaning because of differences in the cultural referents of a concept or word. For example, one item in level C asks students for the synonym of "happy." The correct response is "gay." However, most of the English-speaking students selected other responses because they were familiar only with the new, colloquial meaning of "gay." The students taking the Spanish equivalent of this test selected the correct response because "gay" (feliz, alegre) in Spanish does not connote "homosexual." In this instance, a correctly translated item is culturally biased against English speakers. Item 6, level 2, provides an example of the impact regionalisms have in the Spanish version of an item.

S: choose a gift		accept	aceptar	elegir un regalo
	R	select	R escoger	
		clutch	PD agarrar	
		offer	ofrecer	

Here, "clutch" has been translated as "agarrar." In some regions, "agarrar"

also means to take or choose. Hence, "agarrar" could also be a correct response for this item; this word is a popular distractor for its Spanish version only. Alternative translations of "clutch" are: arrebatar or empuñar. These synonyms do not carry the added meaning of "to choose," hence, they would be good replacements for the popular distractor.

Another translation problem occurs when grammatical forms either do not have equivalents, or else have many of them in another language. For example, a verb which is reflexive or transitive in one language might not have the same characteristic in another language. If the tests are to be as parallel as possible, then form as well as meaning should be considered, particularly where item construction is concerned. Item 12 illustrates this point.

Item 12, level 2

S:	Bothered his brother	Molesto a su hermano
	accompanied	acompañó
	hit	le pegó (hit him)
	irritated	irritó
	told	le dijo (told him)

In this item, the popular distractor, "le pegó," is an attractive response in Spanish for two reasons: first of all, hitting someone is a way of irritating a person; secondly, and perhaps more importantly, this distractor includes the reflexive pronoun "le" which narrows the relationships between "hit" and the stimulus. An analogous situation would be if the English item had been written in the following manner:

Stimulus: Bothered his brother

accompanied
hit him
irritated
told him

If the student has a good idea of the meaning of the stimulus (bothered), then the choice would be either "hit him" or "irritated." Since the test is asking for the "word that means the same or about the same," the student would then have

to consider the role of the pronoun "him." In Spanish, this task would be more difficult since the infinitive of the verb includes the pronoun (pegarle). This problem can be avoided by dropping the pronouns in the distractors or by changing the popular distractor to something less difficult. As it is, this item tests grammar and word usage as well as vocabulary in Spanish, but not in English. Thus, the constructs differ across languages.

Perhaps one of the most difficult features to translate from one language to another is the syntactical style. The translation must reflect the meaning, intent, tone, and general style of the original English version. Yet, the syntactical style must reflect that of the Spanish language. It is not uncommon to find translations which have carried over the syntactical style of the original language. This results in awkward, sometimes confusing language. We found only one instance in each test level where this occurred. For example, in passage 2 of level C, one classroom invites another to a play they are presenting. The following sentence appears in the passage:

> Miss White made a surprise for us so everyone will have cookies after the show. Then it will be time to go home.

Question number 4 asks

> What is Miss White's surprise?
>
> 1. cookies
> 2. going home
> 3. two o'clock Friday
> 4. The Three Little Pigs

The correct answer, number 1 (cookies), was selected by most of the English dominant subjects. However, the Spanish dominant subjects selected number 4 as the popular distractor. The manner in which this sentence was translated seems to be confusing for two reasons. First, the conjunction "so" was translated into Spanish as "y," which really means "and." This transforms the Spanish sentence into a compound sentence where cookies and surprise are unrelated. In addition, distractor number 4 is in bold type in the Spanish version but not in the English, thus making it a more popular distractor in Spanish. Preliminary examinations of

of the biased items suggested that although the quality of the translation is a strong source of bias in some cases, other elements contribute to the problem. Hence, two other potential sources of bias were investigated: curricular relevance and cultural interference.

Curricular relevance. Knowledge of particular concepts, content, and vocabulary can spring from many sources. One source is the material used for reading instruction, such as basal readers. Although basal readers are only a part of the total curricula to which students are exposed, and a small part of their total information sources, they do represent the kinds of content and vocabulary which the students were studying prior to and at the time of testing.

The curricular relevance of the biased items was examined in two ways. First, all key words in the vocabulary and passage comprehension subtests of both levels and languages of the test were counted and compared to the basal readers used for instruction for the sample students (in Spanish and English). Second, the objectives and content of the passages in the basal readers were also searched and examined. The genres in which the passages were written were examined as well as the types of tasks elicited by the test questions. By doing these two analyses, we could approximate the quantity and nature of the presentation of the biased items' contents, at least within the framework of the basal readers.

Comparisons of the test items' vocabulary and contents with those found in the basal readers used for instruction by the sample were based on the texts listed below.

English	Spanish
Keys to Reading, levels: 1-6	Spanish Reading Keys, levels: 1-6
Bolar, levels: 1-3	Bolar, levels: 1-3
Santillana, Hopscotch and Infinity	Santillana, Rayuela and Adelante

To assess the curricular relevance of vocabulary, the frequency with which key test vocabulary appeared in the reading series was counted. Words were described as high frequency if they appeared three or more times within a mean-

ingful context in each volume of a series. Words which appeared less than three times in "meaningful context" or which were used only for phonetics, were labeled "low frequency words." Meaningful context means that the word is presented in such a way that the student is told or can easily infer its definition and function.

Table 2 below shows, in percentages, the amount of words in the tests which also appeared in the basal readers with a high or low frequency of occurrence.

Percentage Ranges of the Match between
the Subtests and Basal Readers on Vocabulary

	LEVEL C		LEVEL 2	
	Spanish	English	Spanish	English
High Frequency	9-34	9-66	12-23	30-41
Low Frequency	7-45	7-25	5-7	19-25

The most significant difference occurs at level 2. Here, the match between the tests and readers is much higher for the English version. The English version also has a higher percentage of high frequency words across both levels of the test. The wide range of most percentages in each cell suggests a marked variability of vocabulary among the readers in both languages.

To assess the curricular relevance of reading comprehension items, three features were examined:

genre, the types of tasks elicited by the questions, and topics.

There were six passages in level C, and 7 in level 2.

The potential effect of genre could not be determined for level C because all of the passages were in the narrative genre. Responses to level 2 items did not cluster around a particular genre either. All of the genres represented in the test were also presented in the basal readers.

Table 3 below lists the types of tasks elicited by the reading compre-

hension questions in both levels of the test.

Table 3

List of Reading Comprehension Tasks Elicited by Questions in the Reading Comprehension Subtests of the CTBS

Task Descriptions	Popular Distractor Items	
	Level C	Level 2
Infer main idea of passage		2, 4, 24, 38
Infer main idea of paragraph	4	22, 43
Infer character's mood/qualities		17, 19, 32,39
Interpret meaning of figurative language		1, 14, 15, 18, 40, 45
Find explicitly stated information	1, 5, 6, 13, 14	13, 21, 41
Derive word meaning from context		7, 9, 28
Sequence of events, facts		25, 29, 34, 35

The most difficult type of questions were those that require the students to infer the main idea, a character's feelings, or the meaning of a metaphor. The next most difficult task was determining the sequences of events or facts and those which require the derivation of word meaning from the text. The least difficult questions are those which ask for information which is explicitly stated in the passage.

All of the basal readers reviewed for the analysis provided a substantial number of opportunities to read, practice, and apply all of the concepts listed in the table. By "substantial number of opportunities" we mean that these objectives were among the core concepts presented by the readers in the lesson plans and reading texts they provide. Therefore, it is likely that the cause of discrepant performances on the biased items arises from sources other than student lack of familiarity with these skills. This would suggest that item problems are due to students' lack of

familiarity with the vocabulary and/or content, assuming that the translation and item construction are sound. Tables 4 and 5 on the following pages list some of the key content examined by the reading passages in levels C and 2. The starred items indicate which topics received specific or related coverage in the readers. In level C, 32 percent of the content and in level 2, 26 percent of the content was also covered in the readers. Of the 12 general topics examined by these tests, six received some amount of coverage in the readers. Only two of these topics, space travel and threshing wheat, were presented similarly in both the test and the readers. There were approximately 70 subtopics listed for these passages; about 26 received some attention in the readers. Only about 8 of them provided matching information for the test and the reader. By matching information, we mean that most of the information provided in the test passage was also provided in the basal readers.

Sometimes there were important differences between the presentations of a topic in the test and in the readers. For example, one of the test topics dealt with the replacement of cut trees to prevent soil erosion and damage to wildlife. Although there were several essays and stories about conservation throughout the reading series, none dealt specifically with the subtopics and information presented in the test. Hence, the basal readers provided only a partial framework and a small portion, if any, of the vocabulary necessary to comprehend the passage. This is corroborated by the number of low frequency vocabulary words in this passage. In some cases, such as with the concept of "farm," the information provided by the readers differs from that presented in the test. The farm, as presented passage 6, level 2 in the test, is a wheat farm of unspecified size. The passage describes how the narrator's grandfather used to thresh wheat, with the help of simple tools, family, and workers. The test passage discusses methods from the past where simple tools were used. The "farms" presented

in the basal readers range from Argentinian and American cattle ranches to a wheat farm in the Midwest. In the texts, the passages were accompanied by illustrations showing a large expanse of land with animals, large machinery, buildings, and many workers. Although some texts used the farm as a setting for stories, the passages mentioned previously were the only ones which described what a farm or ranch is. The article which explains threshing wheat (in one text), discusses modern methods using large machinery.

Theories and experiments related to the importance of a reader's knowledge of a subject upon reading about it have been reviewed by Spiro (1980), Andersen (1977), and others. Background knowledge provides the reader with relevant vocabulary (Shiffrin & Anderson, 1977), syntax (Huggins & Adams, 1980), and concepts (Bruce, 1980; Clark, 1977). All of this information serves as a roadmap of the general structure and content of the passage involved and enhances the reader's comprehension and retention (Rumelhart, 1977; Spiro, 1977). Conversely, the less a reader knows about the subject or content of the reading passage, the harder it will be to grasp essential information and retain it. The poor student response to some of the comprehension items suggests that the lack of compatibility between the content examined by the test and the content presented in the basal readers may be a strong source of bias. Some of the biased items do not have translation problems and examine skills to which the students have been exposed, thus content relevance or cultural factors seem to be the only other sources of bias. The problem of content relevance is complicated when cultural factors interfere.

Cultural interference. This paper cannot present an in-depth discussion of the cultural similarities and differences between Spanish and English speakers. Such a discussion would entail presentation of theories and findings from several disciplines involved in cross cultural studies, such as sociology, anthropology, psychology, among others (cf. Daisen). However,

A When I was a little boy on my grandfather's farm in Kansas, wheat was threshed with simple tools instead of the machines used today.

B When the wheat was brought in from the fields, the hard clay threshing floor was swept clean. Two oxen were hitched to a heavy stone roller. As they pulled, the roller turned in a circle, pressing the wheat hard against the floor to break the seeds from the stems and loosen the husks. (A husk is the covering of a seed. When it falls off the seed, it is called chaff.) Then we picked up bunches of matted straw with homemade pitchforks made from tree branches, which were split into prongs at the end and sharpened to a point. We shook the straw hard until all the wheat seeds were free.

C Next we removed all the straw from the threshing floor, except for about three inches of wheat seeds and chaff. To get this ready for the final process of separating the wheat seeds from the chaff, we used homemade brooms. At the end of a four-foot stick, Grandfather had tied a circular bunch of twigs cut off evenly at the bottom. We swept all of the wheat into several big piles. Not so many years before this, the threshers would have taken bunches of wheat from these piles and thrown them into the air. The wind would have blown the chaff away, and the wheat seeds would have fallen again on the threshing floor. But Grandfather had made a little cart with a sort of paddle wheel in it which made a strong breeze when one turned its crank. As I turned the crank, men shoveled the wheat and chaff in front of the paddle wheel. The chaff was blown off the threshing floor and the wheat seeds fell into a basket.

D When the basket was full, the seeds were poured into bags. By the end of the day all of the wheat was tightly tied in bags, loaded on a cart, and ready for market.

33 What was the surface of the threshing floor like?

1 hard and clean
2 soft and dusty
3 covered with grass
4 covered with weeds

34 After the wheat was put in the bags, it was

5 ground into flour
6 sold at the market
7 kept to feed the cows
8 saved to plant next year

35 In which paragraph is the final process of separating the wheat seeds from chaff described?

1 A
2 B
3 C
4 D

36 What were the pitchforks made of?

5 iron only
6 wood only
7 iron and steel
8 wood and iron

37 Grandfather's way and older ways of separating the chaff from wheat seeds both required the use of

1 oxen
2 wind
3 rollers
4 shovels

38 Which of the following is the best title for this story?

5 "Mowing Wheat"
6 "Planting Wheat"
7 "Growing Wheat"
8 "Threshing Wheat"

39 What kind of man do you think Grandfather was?

1 unkind and lazy
2 thrifty and clever
3 happy and carefree
4 unpleasant to work for

a brief sketch of a few salient cultural differences, with respect to a biased test item, should hint at their effect on students' comprehension of and performance on such an item. Question 39 of passage 6, alluded to previously, which presents the topic of threshing wheat, offers a classic example of bias stemming from cultural interference.

In the passage, the narrator describes how wheat was threshed on his grandfather's farm when the narrator was a child. There is no information about the grandfather other than a description of how he invented a tool and that he owned a farm. The subjects are asked, in question 39:

s:	What kind of a man do you think Grandfather was?	Que clase de hombre crees tu que era el abuelo?
	1. unkind and lazy	1. poco amable y flojo
R	2. thrifty and clever	R 2. frugal e ingenioso
PD	3. happy and carefree	PD 3. feliz y despreocupado
	4. unpleasant to work for	4. dificil de trabajar con el

The majority of subjects taking the Spanish test selected distractor 3 as the correct response. Subjects taking the English test "correctly" chose response number 2. Both groups had received similar coverage of the content in the basal readers, the vocabulary frequency was about the same in both languages, and the translation is very good. Thus, it would be reasonable to conclude that neither the translation quality nor curricular relevance are the major source of bias for this item. The problem lies in the different views the English-and Spanish-speaking subjects may have on the key concepts needed to correctly answer question 39: the farm, thrift, cleverness, happiness, being carefree, and grandfather.

To some Americans, a wheat farm in Kansas probably brings to mind particular notions about grandfather's status and the farm's appearance. Hints of what these notions might be are offered by American literature, history, and the media, especially television. These notions may associate farming with thrift or cleverness.

For the Spanish speaker, the concept of farm, or "granja," as it was

translated, evokes a different scheme. In some countries a "granja" is a chicken ranch, in others it is a more generic and vague term. The relationship between the farm and the work ethic would be based on varying Hispanic work ethics, rather than on the American work ethic. The readers' conception of the grandfather's status are probably quite different.

Student performance was about the same for both groups on those questions which elicited information which was directly offered or implied by the passage. However, on one question, which called for a broader sphere of information for the correct response, the majority of Spanish speakers selected a popular distractor response over the correct response.

CONCLUSIONS

The CTBS provides evidence of three possible sources of bias: problems inherent in the translation; the match between the test and instructional material; and intervening cultural variables. Each of these potential sources of bias can affect the meaning and functions of single words, sentences, and passages, the content of the items, and the skills measured by the item. The degree and manner in which these item features are changed when an item is translated from one language to another will determine whether item equivalence, in all its guises, is maintained. Changes in any of these item features may alter the difficulty or construct of an item. For example, if an item's objective is to measure reading comprehension by asking a question eliciting an inference, the inclusion of inappropriately difficult vocabulary may change the task to one of vocabulary recognition or derivation.

Dual language tests also raise the question of whether, and how, standardized achievement tests can match the wide variety of curricula used by American schools. The general argument, confirmed by research such as the Anchor Study, is that standardized tests vary in their compatibility with the vocabulary, topics, and objectives presented in various

reading materials used for instruction.

The percentage range showing the match between Spanish test items (on the CTBS) and Spanish readers on the dimension of vocabulary (5-41%) suggests a similar pattern for Spanish language standardized tests. However, the test developer states that the content of the items is the same (or nearly so) across languages. This suggests an underlying assumption that curricula are similar or the same for bilingual and monolingual English programs. But the definition and purpose of these two types of curricula would contradict this assumption. The monolingual English curricula assume that all students are proficient in English and thus they concentrate instruction on basic skills or other areas. The bilingual curricula focus instruction on the teaching of English as a second language, teaching other subject areas using a combination of the primary language and English (cf. Spolsky, 1976). This basic difference is manifest in the instructional materials used for both types of programs. One could postulate that if each language version of a reading comprehension item, for example, were altered to enhance its relevance to the corresponding basal reading text (corresponding in terms of language and level), the items may be very different in some of the text features described earlier.

The discussion of cultural interference includes examples which do not reflect the features which have been traditionally assigned to the notion of cultural bias. The passage was translated accurately and contained no improper language or stereotypical character portrayal. All of the biased test items in CTBS, however, indicated the writer's assumption about the intended audience; that is, that the Spanish language reader would perceive the same implied values from the passage as would the English-speaking reader. The fault lies not with the passage but with the question because it elicits knowledge which is external to

to the passage and varies from culture to culture. Thus, culture interferes here not with superficial features of the item but with the assumptions underlying the test question. This kind of problem suggests there may be an additional form of cultural bias.

One definition of cultural interference within the context of reading comprehension in testing could be that the divergent interpretation of a passage (by two cultural groups) is caused by the interjection of one group's cultural attributes into one or more features of that passage. This does not necessarily mean that one group will not comprehend the passage simply because of cultural differences. But it does postulate that one group could interpret some passage features differently. Whether that different interpretation is judged to be correct or not would depend on the test question and the "correctness" of its distractors, as deemed by the developer.

Examination of test bias sources illustrates some of the difficulties which plague "equivalent" dual language achievement tests. It also shed some light on a few of the assumptions some educators have regarding these instruments. These assumptions concern the equivalence of two language versions of the same test on the features of language, content, format, difficulty, and curricular match.

The CTBS and its Spanish version are, for the most part, equivalent in terms of vocabulary, content, and format. The Spanish language test is relatively free of language which might favor one ethnic group over another. The translation is generally accurate and the format is identical across tests.

However, examination of curricular match in terms of vocabulary and general topics suggests that the English language version has a stronger match to English basal readers. Since the content and vocabulary is the same across language versions, the tests' closer match to English basal

readers may reflect the fact that monolingual English and bilingual program curricula are probably different in terms of vocabulary and content.

The problems of cultural interference which were discussed earlier suggest that there may be subtler, more elusive forms of cultural bias such as the interjection of values or associations which reflect one culture and not another.

These last two bias sources suggest that the intended "audience" or test taker, as well as the tests themselves, should be a consideration in test bias. The original English version of CTBS was written for test takers who were proficient in English and receiving a monolingual English curriculum. This type of test taker may differ from the monolingual Spanish speaker in terms of test wiseness, curricular history, and cultural associations regarding several aspects of the test, ranging from its vocabulary to the student's perception of achievement testing. Spanish dominant speakers are given tests in their primary language because they are not proficient in English and, by extension, are not familiar with many concepts taken for granted by Americans. There may be differences between Spanish- and English-speaking test takers beyond those of language which influence their perceptions of and responses to test items.

The findings of this analysis provide strong evidence suggesting that translation, curricular match, and cultural interference are the causes of bias. However, further empirical studies are necessary to isolate the cause of bias for individual test items, whether in CTBS or any other test. Second, studies providing more information about biases stemming from cultural interference and curricular match are needed in order to develop a methodology, beyond the statistical approaches we now have, which will provide us with a fuller understanding of the issue.

Table 4

List of Topics for Level C, CTBS

Passage Number	General Concept	Subtopics
1	*Vacation on the beach	letters sailboats* uncles sea shells sand dogs
2	A school play*	letter/invitation classroom visit to other classes classroom party a surprise for the classroom time (i.e., 2 o;ckicj) The Three Little Pigs teachers*
3	Animal rescue	horses* personification of animals* mountain lions rescue by animals rescue by people (cowboys) cowboys*
4	Birthday surprise	box of candy as a gift toychest going to school* personification of dog* mischieviousness of dog/pet
5	New shoes	going to school with new shoes teacher reaction to this character's emotion
6	Space travel	astronauts* spaceship* traveling to the moon* change in perspective as you get nearer to the moon and farther from earth

Table 5

List of Topics for Level 2, CTBS

Passage Number	General Concept	Subtopics
1	Skyscrapers being tired & lonely	personification of inanimate objects metaphors* loneliness* tiredness*
2	Conservation of forests	replacing cut tress with seedlings erosion preventing erosion economical damage of erosion fisheries salmon spawning erosion damage to spawning beds
3	Coming home from a camping trip	camping* child camping without family learning new things when camping homesickness the city as a 'home'* skyscrapers at night bus trip*
4	The abalone	abalone's feeding habits, appearance, and habitat suction cup abalone as a food product abalone as an inedible product
5	Neil Armstrong steps on the moon*	Armstrong's professional history flashbacks the Korean war the X-15 spacecraft* astronauts* stepping on the moon as: awe inspiring & a new phase for mankind
6	*Threshing wheat	*farm grandfather Kansas *simple tools vs. machines *description of specific tools: stone roller, pitchfork, prongs, thresher, crank, paddle wheel *inventing a tool taking the wheat to market

Table 5 (continued)

Passage Number	General Concept	Subtopics
7	Jet-plane ride	*birds-eye view from a plane *comparison of people & land to patches and ants *perspective of view from plane as it ascends *feeling static when actually moving *"knowing with the top of our minds" "solid air" beneath the plane the plane as a microcosm or little world forgetting that microcosm when landing and returning to daily life

BIBLIOGRAPHY

Adams, M. J. Failures to Comprehend and Levels of Processing in Reading. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), Theoretical Issues in Reading Comprehension. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1980.

Adelante, Guía Para el Profesor. Santillana Publishing Company, 1977.

Anderson, R. C. The Notion of Schemata and the Educational Enterprise: General Discussion of the Conference. In R. C. Anderson, R. J. Spiro, & B.E. Montague (Eds.), Schooling and the Acquisition of Knowledge. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1980.

Anderson, R. C. & Shiffrin, Z. The Meaning of Words in Context. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), Theoretical Issues in Reading Comprehension. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1980.

Bruce, B. C. Plans and Social Actions. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), Theoretical Issues in Reading Comprehension. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1980.

Cummins, J. Linguistic Interdependence and the Educational Development of Bilingual Children. Review of Educational Research, Spring, 1979, 49(2), 222-251.

Doehring, D. G. & Aulls, M. W. The Interactive Nature of Reading Acquisition. Journal of Reading Behavior, 1979, 11(7), 27-40.

Harris, T. L., Creekmore, M., Matteoni, L., & Allen, H. B. Air Pudding and Wind Sauce, Teacher's Manual. Oklahoma City, Ok.: The Economy Company, 1975.

Harris, T. L., Creekmore, M., Matteoni, L., & Allen, H. B. Mysterious Wisteria, Teacher's Manual. Oklahoma City, Ok.: The Economy Company, 1975.

Huggins, H. W. F. & Adams, M. J. Syntactic Aspects of Reading Comprehension. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), Theoretical Issues in Reading Comprehension. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1980.

McArthur, D. Detection of Item Bias Using Analyses of Response Patterns. Los Angeles, Ca.: Center for the Study of Evaluation, University of California, Los Angeles, 1981.

Rayvela Guía Para el Profesor. New York: Santillana, 1977.

Rodriguez-Bov, I. A Study of the Parallelism of English and Spanish Vocabularies. Puerto Rico: Superior Educational Council of Puerto Rico, 1950.

Rumelhart, D.E. Schemata: the Building Blocks of Cognition. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), Theoretical Issues in Reading Comprehension: Perspectives and Cognitive Psychology, Linguistics, and Artificial Intelligence and Education. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1980.

Saville-Troike, M. R. Bilingual Children: a Resource Document. Arlington, Va.: Center for Applied Linguistics, 1975. (ERIC Document Reproduction Service No. 102867)

Schmitt, C. J. A Cada Paso (Teacher's ed.). New York: Webster Division, McGraw-Hill Book Company, 1978.

Spiro, R. J. Remembering Information from Text: the State of Schema Approach. In R. C. Anderson, R. J. Spiro, & W. E. Montague, Schooling and the Acquisition of Knowledge. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Spiro, R. J., Bruce, B. C., & Brewer, W. F. (Eds.). Theoretical Issues in Reading Comprehension: Perspectives and Cognitive Psychology, Linguistics, Artificial Intelligence, and Education. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1980.

Spolsky, V., et al. A Model for the Description and Perhaps Evaluation of Bilingual Education. Navajo Reading Study Project Report (Vol. 23). Albuquerque, N.M.: New Mexico University, 1974.

Trouble and Turnips, Reading Basic Plus. New York: Harper & Row, 1977.

U.S. Commission on Civil Rights. A Better Chance to Learn: Bilingual Bicultural Education. Washington, D.C., 1975.